DOCUMENT RESUME

ED 249 996 IR 050 864

AUTHOR Turock, Betty J.; Shelton, Hildred C.
TITLE Subject Access in the Small Library.
PUB DATE May 84
NOTE 13p.; Paper presented at the Mid-Year Meeting of the American Society for Information Science (Bloomington, IN, May 20-23, 1984).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Information Needs; *Information Retrieval; *Library Catalogs; Library Research; Microcomputers; Online Systems; Public Libraries; *Search Strategies; *Subject Index Terms; Use Studies; Vocabulary Development
IDENTIFIERS Library of Congress Subject Headings; *Library Users; Online Catalogs

ABSTRACT

This pilot study on the subject access problems of patrons of small to medium size libraries was designed to measure the extent to which users' vocabularies matched the search vocabulary of bibliographic records in the card catalog, and to enhance subject access by developing a microcomputer system which integrated Library of Congress Subject Headings (LCSH) with the natural language of the users. Three public libraries in Virginia were selected as test sites because their users encompassed a heterogeneous population that cut across demographic and socioeconomic indicators. Data collected on exception cases (times when users asked librarians for assistance in finding information) revealed that of the 412 questions analyzed, almost 60% produced no match between the language of the information seekers and the information organizers. The online system which was developed allows directed browsing and puts descriptors into context. It contains eight fields: Subject, the primary access point containing the users' vocabulary; Enter, which serves a cross-referencing function; LC, a selection of relevant Library of Congress subject headings; Broader, Narrower, and Related, three fields which supply hierarchically oriented connections; Scope Notes, which define the meaning of records in the Subject field; and Bibliography, which lists other pertinent holdings. (Author/THC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SUBJECT ACCESS IN THE SMALL LIBRARY

Betty J. Turock
Rutgers University
School of Communication, Information and Library Studies (SCILS)
185 College Avenue
New Brunswick, NJ 08903
Hildred C. Shelton
Pittsylvania County Public Library
South Main Street
Chatham, VA 24531

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Betty J. Turock
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED249996

ABSTRACT

A pilot study, recently conducted in Virginia, concentrated on the heretofore virtually ignored subject access problems of the small to medium size library. The research objectives were: 1) To measure the extent users' vocabularies matched the search vocabularies of bibliographic records found in the traditional card catalog, and 2) To enhance subject access by developing a microcomputer driven system which integrated Library of Congress Subject Headings (LCSH) with the natural language of users.

Three public libraries - Pittsylvania County, Roanoke County and Danville - served as the test sites. Public libraries were selected, since their users encompass a heterogeneous population that cuts across demographic and socio-economic indicators. Data collected on exception cases, that is, times when users asked librarians for assistance in finding information, revealed that of the 412 questions analyzed, close to 60 percent produced no match between the language of the information seekers and the information organizers. The exception cases were identified for enhanced access, since they recorded subject areas useful to current library clientele but difficult for them to retrieve.

The online system developed allows directed browsing and puts descriptors into context. It contains eight fields. The Subject field, the primary access point, contains the user's vocabulary. The Enter field serves a cross referencing function. Broader, Narrower and Related field supply hierarchically oriented connections. The Scope Note defines the range and meaning of the subject. Other pertinent holdings are listed in the Bibliography field. The system, which permits truncated searching, is available at low cost.

IR052864

INTRODUCTION

The myth of the known item search was put to rest in a study of online catalog use conducted by the Council on Library Resources. Through the investigation it was determined that the major portion of users' searches were subject orientated explorations for unknown items.[1] The June, 1982 meeting devoted to the topic, sponsored by the Council on Library Resources, added further credence to the fact that subject access demands consideration as one of the major issues facing libraries of all types now and in the future.[2]

To date, attempts to resolve problems in subject access have focused on the large research institutions with minimal attention given the small to medium size organization. This is not surprising, since the early computer technology available to address subject access was limited to large mainframe systems or minicomputers often beyond the financial reach of the smaller library. With the advent of the microcomputer new opportunities are available to facilitate subject searches at low cost. Among them is the ability to develop systems capable of providing a mediating link between the natural language of information seekers and the precordinate controlled vocabularies employed by information organizers.

THE PROBLEM

Prior to attempting to develop a system capable of enhancing subject access in the small to medium sized library, it was necessary to ascertain the level of access currently afforded users through searches conducted via the Library of Congress Subject Headings (LCSH) found in the traditional card catalog. Intuitive contentions that users' search terms frequently have nothing in common with LCSH surfaced repeatedly in the professional literature. Attempts to locate evaluations of those contentions, however, led to the conclusion that very little had been reported concerning the operation of the smaller library in this regard. Within the public library sphere alone, 82 percent, or more than 12,000 of the total 14,831 libraries, serve populations under 25,000.[3] When small to medium academic, special and school libraries are added, the number is further inflated and the lack of reliable information on which to develop systems becomes more ironic.[4] In fact, it is apparent that the problems of subject access in the smaller library are virtually unknown beyond those information professionals directly involved with them.

CREATING THE FRAMEWORK

The ongoing objectives of the work reported here have been to: (1) Identify and measure the extent users' vocabularies match the search vocabularies of the bibliographic records in card catalogs, (2) Create a retrieval format which integrates LCSH with users' natural language search vocabularies, (3) Develop a

low-cost, online system to enhance subject access, and (4) to compare the success rate of searches done on the traditional system with that of the augmented system. One to three have been completed in the pilot study reported on here. Four remains for the future when there is sufficient information in and on the system to make such a test valid.

For the purpose of this investigation, subject access was defined according to Butler as "the set of processes and techniques used in the representation of a work so that its contents may become known to one desiring the information therein without prior knowledge of the existence of the work, its authorship, or location."[5]

To determine the extent of match between users' vocabularies and the LCSH search vocabulary, exception cases were recorded and analyzed. Exception cases were defined as instances when users requested help from librarians in finding information. If users did not seek assistance, it was assumed that their search vocabularies and the vocabulary of the LCSH coincided. Several further assumptions were made, including that:

1. Small libraries are likely to use LC cataloging as is.

2. The LCSH will be the basis from which most smaller libraries will initiate online public catalogs or modified versions of such catalogs as they endeavor to increase subject access. In fact, one of the major weaknesses of current online catalogs is that they are too frequently merely automated versions of Library of Congress catalog cards.

The pilot was conducted in three public libraries in Pittsylvania County, Danville and Roanoke, Virginia. They were selected as the test site, since their users encompass a heterogeneous population, one that cuts across demographic and socioeconomic indicators with Pittsylvania, Danville and Roanoke found in ascending order in most instances. The three vary from small to medium in size with the highest budget significantly less than one million dollars; they were located in rural, town and urban settings with differing population profiles. Table 1 provides an overview of the sites.

The libraries within the three communities were equally varied as Table II attests.

Data were collected over a period of two weeks on a time sampling basis in October, 1983 and again in January, 1984. All questions recorded for analysis were exception cases. No known item searches were included. If, upon perusing the LCSH, a variation of the term sought was found in the catalog, that search was denoted as a match (M) between the user's vocabulary and the LCSH. For example, if information requested on diets was found under DIETING, it was tabulated as an M. When, however,

Table I. Profile of the Virginia Test Sites[6]

Library	Population Served	Population Density Per Square Mile Served	Median Household Income	Median Years of School Completed by Persons 25 Years and Over	Percent of Persons 25 Years and Over Who Are High School Graduates	Percent of Persons 25 Years and Over Who Are College Graduates
Pittsylvania County Public Library	66,147	66	$14,020	10.4	37.8	5.7
Danville Public Library	45,642	268	13,413	11.7	47.5	10.5
Roanoke County Public Library	72,945	290	20,205	12.6	70.0	17.7

BEST COPY AVAILABLE

Table II. Profile of the Three Libraries in the Pilot Study[7]

Library	Full-Time Equivalent Staff Positions	Professional Positions	Total Budget*	Circulation*	Volumes* Held	Circulation* Per Volume
Pittsylvania	10.75	2	$184,115	105,942	47,877	2.2
Danville	16.7	5	$339,966	193,019	83,918	2.3
Roanoke	33.5	8	$643,778	654,685	209,326	3.1

*Figures are taken from the 1982 fiscal year.

BEST COPY AVAILABLE

users' terms and LCSH lacked a similar nomenclature, that search was denoted as no match (N). For example, if information requested on Gun Control was found under FIREARMS - LAW AND LEGISLATION, it was calculated as an N. Finally, data were compiled on exception cases, designated N, which required the formulation of a sophistication search strategy (S). For example, a request for a book on Song Writing, located under a SEE reference pointing to the LCSH Music, Popular Writing, required no strategy. However, information sought on the effect of discontinuing food subsidies at a local hospital on the nutritiousness of student nurses' diets required a more complex approach and was denoted as an S. Table III supplies the results of the subject searches in the exception cases.

There were 412 total exception cases recorded. In 170, or 41%, of them there was a match between LCSH and the users' vocabulary and in 242, or 5 %, there was no such match. Danville and Pittsylvania County had their highest tallies among the exception cases in the no match category. For them the next highest figures were matches and the third were questions requiring search strategies. In Roanoke, however, questions requiring search strategies came in first, no match second and matches third. In Roanoke, then, the information requested required a higher proportioning of sophisticated search strategies, or simply, more sophisticated users asked more sophisticated questions.

Table III. Subject Search Analysis

Libraries	Total Questions Asked	Match Between LCSH and User Vocabulary	No Match Between LCSH and User Vocabulary	Questions Requiring Search Strategies
Pittsylvania	56	20	36	2
	.14	.36	.64	.06
Danville	316	142	174	12
	.76	.45	.55	.07
Roanoke	40	8	32	21
	.10	.20	.80	.66
Total	412	170	242	35
	1.00	.41	.59	.14

While data must be collected over a longer period of time and on a national basis for any broad generalizations to be made with assuredness, this study did confirm suspicions that, at least in the three case sites, LCSH and users' vocabularies were sufficiently at variance to initiate the development of a system which would enhance subject access. As a result, the subset of holdings denoted as exception cases were identified for enhanced access in the system design phase, since such cases pointed up subject areas useful to current library users but difficult for them to retrieve.

ENHANCING SUBJECT ACCESS

The objective of the system developed to enhance subject access is to integrate LCSH with users' search vocabularies as well as with the search vocabularies of other sources, such as the Reader's Guide, and to create a hierarchy of the resulting descriptors. A thesaurus/bibliography was developed and stored at the Pittsylvania County Public Library on an Apple II microcomputer with two 5-1/4" floppy disk drives. The software is PFS: Files by Software Publishing Corporation, which permits the format of the files to be established, then input, search/edit, print and delete modes to be activated. The hardware and software are currently available for approximately $2,200.[8]

The data structure developed to meet the objective was conceptualized as containing eight fields, electronically displayed as shown in Figure 1.

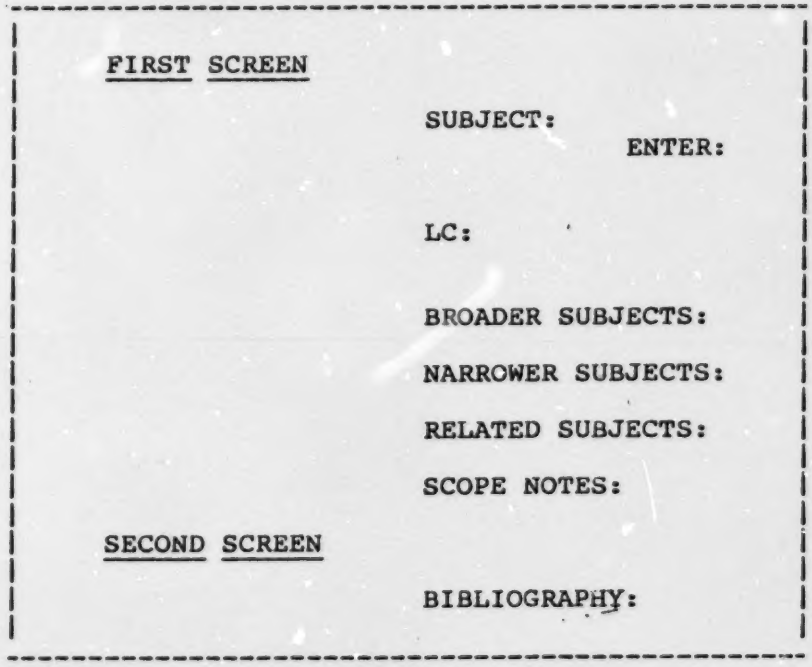

Figure 1. Screen Display of the Thesaurus/ Bibliography Structure

The SUBJECT: field, the primary access point, contains words or phrases taken from the vocabularies of users when they submitted requests for information. To determine the words and phrases to include, actual reference questions were collected and analyzed. As an example, the descriptor CHILD ABUSE was located in the subject field when it was determined that users were more apt to employ it than the Library of Congress headings CHILDREN (CARE AND HYGIENE) or CHILD PSYCHOLOGY.

Data in other fields help users find the specific materials they are seeking. The ENTER: field instructs the user to register another term in order to retrieve a record containing a bibliography. This field is also used as a cross-referencing device, linking a term not used as a primary access point to one that is. The term may be another word commonly used by patrons, as MADD for Mothers Against Drunk Driving. The field can assist the user in another way. If the word or phrase in the SUBJECT: field represents a complex idea or concept, the user is instructed to scan the BROADER, NARROWER and RELATED SUBJECT: fields in which several other terms are offered. The user may then enter these terms in order to find the bibliography that will have the most appropriate material.

A selection of Library of Congress Subject Headings that most nearly match the commonly used term are entered in the LC: field. The intent is to link the user's language with subject headings found in the library's card catalog, that is, to connect the user's natural language and the LCSH. This also makes it possible for the user to return to the card catalog and find more information, if it is desired.

To suggest the hierarchical structure of the subject organization, the next three fields, BROADER SUBJECT:, NARROWER SUBJECT: and RELATED SUBJECTS: are developed. They also lead the user to more material on the topic. The BROADER SUBJECT: field transfers the user to concepts that are more general than those in the SUBJECT: field. The user may reenter one of these concepts when shown with the symbol /X// as the broader subject WASTE under WASTE DISPOSAL. The words and phrases in the NARROWER SUBJECT: provide an opposite function, indicating ideas embodied in the terms in the SUBJECT: field, such as the narrower subject HERPES under VENERAL DISEASE.

The RELATED SUBJECTS: field is used to link terms with similar meanings to the terms in the SUBJECT: field. These terms can be suggested by the wide variety of subject headings under which material on the topic is found. In the case of "HYPNOSIS" information is found under "BEHAVIORAL MODIFICATION," "REINCARNATION" and "BRAINWASHING," so the user is alerted to other perceptions and usages of the term in the SUBJECT: field. A request for myths and legends prompts three separate bibliographies in the database: MYTHS, LEGENDS, and MYTHOLOGY. All are closely related but not identical, so under each term used in the SUBJECT: field, the other two terms appear in the

RELATED SUBJECT: field. In addition to defining the meaning of the term in the SUBJECT: field, the RELATED SUBJECT: field gives an understanding of the scope of the material in the bibliographies. The SCOPE NOTE: field is used to define precisely the meaning of the records found in the SUBJECT: field, as in the records under MYTHS, LEGENDS and MYTHOLOGY.

In the BIBLIOGRAPHY: field the books and periodicals are listed that pertain to the term in the SUBJECT: field. Call numbers, authors, titles and publishing information are given. Precise page numbers for parts of books and periodicals that are available in the library, on the shelf or microfiche, are included. Reading levels are indicated in some instances. When the title does not clearly identify the content of the book or article, a brief statement summarizing its content is added, similar to the scope notes in CIP data.

CREATING THE DATA BASE

After conceptualization of the data structure was completed, the procedures were set for inputting data.

THE REFERENCE QUESTION FORM, shown in Figure 2, was designed for staff members to capture the necessary information from users' requests.

REFERENCE QUESTIONS Date_____

Check the appropriate description of the information seeker:

Adult ____ Young Adult ____ Child ____ Student ____ Other ____

Approximate Reading Level Required:_____

Question as stated by patron:_____

_____

_____

Question restated, if necessary:_____

_____

General subject area used to find information:_____

_____

Specific sources in which information was found:_____

_____

Check here if no information was found: ____ Staff initials: ____

Figure 2. Collecting Data for System Input

First, the books and periodicals listed under "Specific sources in which information was found" are carefully checked for the correct bibliographic information. Each source is also reviewed for its relevance to the term in the SUBJECT: field. Using these sources, a clear understanding of the question is formulated from which the cross-referencing structure is developed.

Second, the subject headings to locate the sources in the card catalog are traced and each checked further for more pertinent material. From the list of subject headings gathered and the formal headings used for similar material by the Library of Congress, several headings are chosen for inclusion in the LC: field of the record on the basis of their compatibility with the natural language terms of the user.

Third, a word or phrase that embodies the intent of the user's query is decided upon; as often as possible the user's exact wording is employed. With this established, the BROADER, NARROWER and RELATED SUBJECT: field structure is built. Information about the source material is gathered from the fly-leaf, table of contents, preforatorial material and CIP data. The need for this framework of cross-referencing proceeded from the intention to place bibliographies under the most specific term. For example, under FROGS which is addressing the specific question of frog reproduction, EMBRYOLOGY is provided as the broader term where explanatory material can be found.

Fourth, the bibliographies are entered. By using the most specific term they are kept short and to the point. Material that treats the idea or concept in a general manner is located under a term referred to in the BROADER SUBJECT: field. The form of the work is indicated where helpful, including formats other than the book.

While developing the cross-referencing structure, the whole universe of possibilities connected with the question is not considered, rather the subjects and terms are limited to those that arise easily from the users' questions. If in the future more detail is required, more terms can be added, and the structure developed further. As this thesaurus structure emerges, it becomes a tool to aid in that development.

EQUIPMENT AND DESIGN

Development of the system was undertaken at the Pittsylvania County Public Library on an Apple II microcomputer with two 5 - 1/4" floppy disc drives. The software selected was PFS: Files by Software Publishing Corporation, which permits the format of the files to be established, then allows input, search/exit, print and delete modes to be activated.

The software also allows for truncation of terms, so the data base can be searched by using words, phrases, and parts of

either when set between double periods (.. ..). Using the truncated forms slows response time, since the whole data base will be searched and some irrelevant records retrieved. The ability to search with truncated terms gives the user and the designer flexibility. For example, users looking for a bibliography under MOTHERS AGAINST DRUNK DRIVING (MADD) can enter the whole phrase if it is remembered exactly, or MOTHERS, DRUNK DRIVING or parts of either. The hardware and software used in this test system are currently available for less than $2,000. The test system was designed for use in the pilot phase of the project only. The extent of the file is 1,000 items of 128 characters. When the system has been tested sufficiently to determine the most desirable format, using a custom written program, the data base will be placed on a Wincester disk which will hold approximately 33,000 entries.

LANGUAGE CONVENTIONS

Abbreviations are not used in retrievable fields, except in the case of familiar or popular acronyms employed by users, such as AA for Alcoholics Anonymous. Full names are cross-referenced to the abbreviation, so that errors in searching caused by the user's inability to remember the exact form of the entry are avoided. Since users do not invert proper names in formulating their questions, they are not inverted in the system. In fact, employing the broader and narrower term structure makes inversion unnecessary. Plural nouns are used in all subject headings. Modifiers are kept to a minimum. Occasionally when a phrase such as "EFFECT OF DIVORCE ON CHILDREN" must be used, they are carefully considered. Staff consultations are held to help isolate the wording most likely to be part of the user's vocabulary.

FOR THE FUTURE

The current system combines a number of features which the Subject Access Conference touted as ideal, including:

1. A means by which LCSH are integrated with other vocabularies and a means to switch between them.

2. A thesaurus screen which offers the user an opportunity to do directed "browsing," while avoiding the inherent order of alphabets.

3. Reduction of false hits because the thesaurus structure puts descriptors into context.

4. Use of class numbers to guide patrons to better search strategies, since the bibliographic components of the record contain call numbers of relevant materials.

5. Inclusion of journal materials in the subject access data bases.[9]

The initial tests have led to modifications in the system. Data is still collected on exception cases. A comparison between the success rates of the standard data base and the augmented data base is projected to determine whether we have made a significant improvement in access or merely created another data base which is equally difficult to negotiate.

REFERENCES

1. Brownrigg, Edward B. and Lynch, Clifford A. "Online Catalogs: Through A Glass Darkly," Information Technology and Libraries 2(1): 104-115; 1983.

2. Subject Access: Report of a Meeting Sponsored by the Council on Library Resources. Washington, D.C.: Council on Library Resources, June, 1982.

3. American Library Directory. New York: R. R. Bowker, 1981: x.

4. Vavrek, Bernard. "A Struggle for Survival: Reference Services in the Small Public Library," Library Journal 108: 966; 1982.

5. Subject Access, p. C-2.

6. U.S. Census, 1983. Virginia Counties and Cities Date Book, 1983. Norfolk, Virginia: William C. Overman, 1983.

7. Statistics of Virginia Public Libraries and Institutional Libraries, 1981-82. Richmond, Virginia: Virginia State Library, 1983.

8. Turock, Betty J. and Shelton, Hildred C. "Online Catalog in the Small Public Library," Proceedings of the Fifth National Online Meeting. Compiled by Martha E. Williams. Medford, New Jersey: Learned Information, 1984. Forthcoming.

9. Subject Access, p. 32.

ACKNOWLEDGMENTS

Turock and Shelton gratefully acknowledge the conceptual and technical assistance of Martha C. Bruning, Pittsylvania County Public Library on the design of the system and the preparation of this paper, as well as the critiques provided by Henry Voos and James Anderson of Rutgers School of Communication, Information and Library Studies.